CHAPTER 2

LITERATURE REVIEW

2.1 Introduction of MOSFET

The structure of the MOS field-effect transistor (MOSFET) has two regions of doping opposite that of the substrate, one at each edge of the MOS structure as shown in **Figure 2.0**.These regions are called the source and drain, and a pn junction exists between them and substrate. When terminals are connected to all the various regions of the MOSFET, a four terminal device results, with the terminal designate as G (gate), S (source), D (drain) and B (substrate).

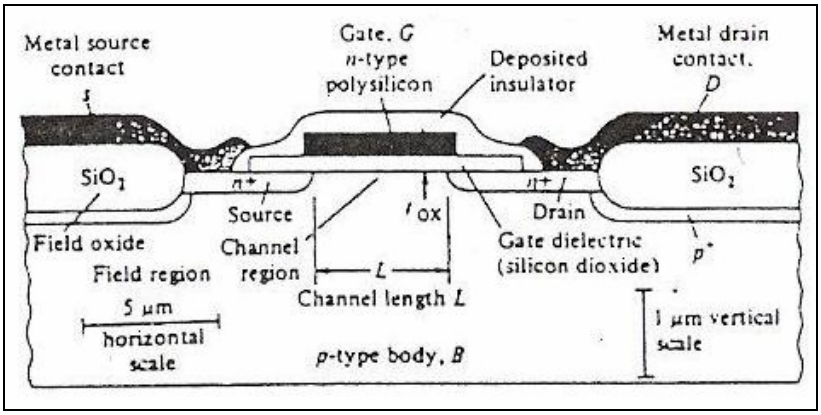

Figure 2.0: Structure of MOS device [8]

Since one of these terminals can be designated as the common terminal, three independent terminal voltages can be applied to the MOSFET. However, only one significant current exists in an ideal MOSFET. That is, we assume that the gate current is zero ($I_G = 0$) and that the source and drain junctions are always kept under reverse bias during normal MOSFET

operation. Since reverse bias current in a pn junction can be considered negligible (and I_G = 0), substrate current will also be inconsequential (I_{sub} = 0).Thus only the drain current I_D which flows between the source and drain in the MOSFET needs to be considered. In summary, three independent terminal voltages and one current, I_D is generally associated with the operation of the ideal MOSFET.

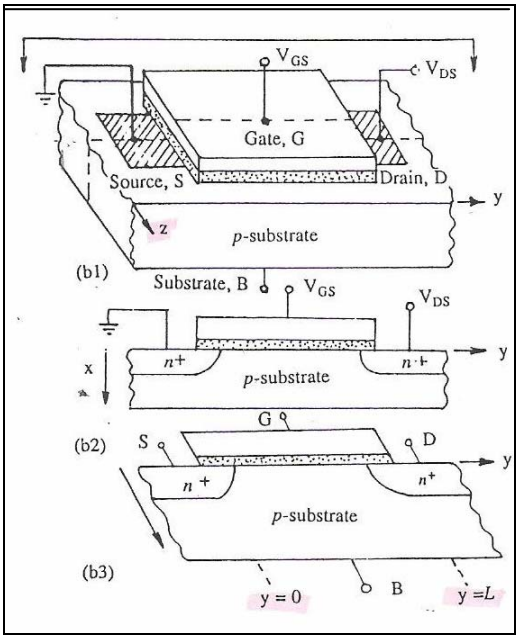

Figure 2.1: (b_1) Perspective view of the MOSFET structure, (b_2) Cross section of the MOSFET cut down the middle of the channel, (b_3) Cross section of the MOSFET lying flat[8]

Figure 2.1 defines the axes to be used in the MOSFET structures. That is, the x-direction is perpendicular to the Si-SiO_2 interface (vertical direction), with x = 0 at the Si surface. The y-direction is parallel to the Si surface in the direction from source to drain (lateral or longitudinal direction).Since the source and drain are separated by a distance L or the channel length, y = 0 at the source end of the channel and y = L at the drain end. The z-direction is the other direction parallel to the Si surface (perpendicular to the y-direction) and defines the channel width.

2.2 MOSFET Device Structure

2.2.1 Basics Operation

The operation of a MOS transistor involves the application of an input voltage to the gate electrode. This establishes an electric field perpendicular to the Si-SiO_2 interface in the channel region of the device. The conductance of the channel region can be modulated by varying this electric field. Since an electric field is responsible for controlling the output current flow, such devices are termed field-effect transistors (FET_S).

As Wolfs [9] explaining that, If no gate bias is applied, the circuit path between source and drain consists of two back-to-back pn junctions in series. In this case, if V_{DS} is applied, I_D will consist of only the reverse-bias diode leakage current, which is normally negligible. When positive bias is applied to an NMOS transistor gate, however, electrons will be attracted to the channel region and holes will be repelled. Once the positive gate voltage become strong enough to form an inversion layer, an n-type channel is formed that connects the source and drain regions.

According to Ren-Ji theory [10], a drain current, I_D can then flow if a voltage V_{DS} is applied between the sources and drain terminals. In the simplest analysis, the voltage-induced n-type channel is assumed that it does not form unless the voltage applied to the gate exceeds the threshold voltage, V_T.

As Hess [11] points out MOS devices in which no conducting channel exists when V_{GS} =0, are referred to as enhancement-mode or normally OFF transistors in **Figure 2.2** and **Figure 2.3**.With NMOS enhancement-mode transistors, a positive gate voltage, V_{GS} greater than V_T must be applied to create the channel or to turn them ON, while to turn on PMOS enhancement-mode devices, a negative gate voltage whose magnitude is $>V_T$, must be applied. Note that in NMOS transistors a positive voltage must also be applied to keep the drain-substrate reversed-biased,

while in PMOS devices this voltage must be negative. On the other hand, it also possible to build MOS devices in which a conducting channel region exists when V_{GS} =0V and such MOS devices are described as being normally ON. Since a bias voltage to the gate electrode is needed to deplete the channel region of majority carriers, and thus turn them OFF, such devices require a negative gate voltage to be turned OFF, while corresponding PMOS devices require positive gate voltage.

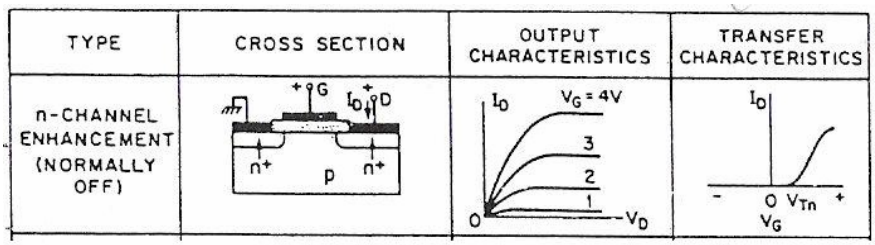

Figure 2.2 : N-channel enhancement mode [8]

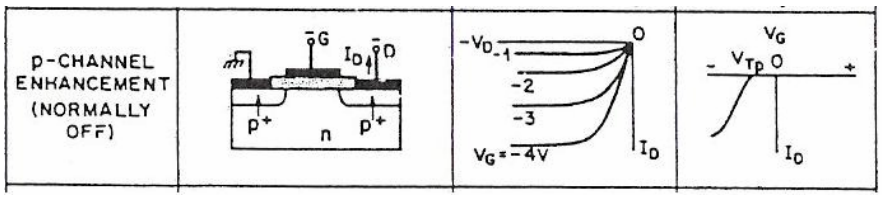

Figure 2.3: P-channel enhancement mode [8]

2.2.2 Biasing the Inversion Layer

An external bias voltage can be applied to the channel region of a MOS structure) if an inversion layer has been induced there. It is called “communication”. On the other hand, if the channel region is not inverted, the application of a bias voltage to a diffused region adjacent to the channel region will have no effect on the surface potential in the channel. Such “communication” has the following impact on the behavior of a MOSFET:

1. If the gate voltage applied to a MOSFET is less than the threshold voltage, the device said to be operating in the subthreshold regime.
2. If the channel is inverted, and channel–bulk voltage, V_{CB} is increased, the depletion region of the field-induced channel increases. If gate-bulk voltage, V_{GB} is held fixed, the number of electrons in the inversion layer will decrease as V_{CB} is increased. At some value of V_{CB} the inversion layer charge will entirely disappear.
3. "Communication" with the channel also permits the diffused region to become an additional source of mobile carriers to the channel.

2.3 Long Channel MOSFET

2.3.1 Circuit Characteristics

From the perspective of the MOSFET as a circuit component, long-channel behavior has been specified in a number of ways. A list of device parameters from a circuit perspective deemed to be characteristics of long-channel MOSFETs include the following:

1. The threshold voltage V_T is independent of channel length L.
2. The threshold voltage V_T is independent of drain bias voltage, V_{DS}.
3. The drain current in saturation I_{Dsat} is independent of V_{DS}.
4. The drain current has a linear dependence on 1/L.
5. The subthreshold current I_{DS}t is independent of drain bias.
6. The subthershold swing St is independent of gate length.

2.4 Threshold Voltage Control in MOSFETs

Qiang Chen.et al [12] in explaining the definition of the threshold voltage is the value of the gate voltage that turns on the transistor by inducing a highly conductive channel from the source to the drain. While, Michael Shur [13] found that depending on the applied gate to source bias, V_{GS}, any field effect transistor can be either in the on-state with a conducting channel between the source and drain, or in the off state, with practically no conduction between the source and drain. The gate voltage, V_T separating these two regimes is called the threshold voltage.

The factors that impact V_T is given in Equation (1).An examination of each term will reveal the device parameters that can be adjusted to provide practical control of V_T.

$$V_T = \varphi ms - Q_{ot}/C_{ox} + 2\sqrt{\kappa_{si}\varepsilon_o q N_{sub}\varphi_B}/C_{ox} + 2\varphi_B \quad \text{..... (1)}$$

The φ_{ms} term depends on work function difference between the gate, $q_{\varphi m}$ (gate), and the semiconductor, $q_{\varphi B}$ (sub).While $q_{\varphi m}$ (gate) for metal and heavily doped silicon gates is constant, the parameter Φ_B (sub) depends on the substrate doping- but only in a logarithmic manner. Hence, each factor of 10 increase in substrate doping will change the φ_{ms} term by only 2.3Kt/q or ~0.06 V (kT / q =0.026V at 300K).Thus, changes in the substrate doping concentration produce changes in V_T through the φ_{ms} term); thus, the $2_{\varphi B}$ term is also ineffective for controlling V_T. Since every attempt is made to keep Q_{tot} as low s possible through various processing procedures and C_{ox} is relatively large (since t_{ox} is very thin in submicron MOSFETs), the Q_{tot}/C_{ox} term is also very small in modern MOSFETs. Hence, this term must also be ruled out as a candidate for controlling V_T. While is true that C_{ox} could be varied (primarily by changing $t_{ox),}$ this is not a practical approach to controlling V_T in active devices, since tox is normally made as thin as possible to maximize I_D. This leaves $2\sqrt{\kappa_{si}\varepsilon_o q N_{sub}\varphi_B}/C_{ox}$ term as the remaining candidate for controlling V_T in active devices. Since the change of V_T in this term depends on $\sqrt{N_{sub},}$ it indicates that V_T control is possible by exploiting this effect.

Merely increasing the substrate doping, however, is not desirable since it will adversely impact other MOSFET characteristics, such as lower junction- breakdown voltages, larger junction capacitance, and lower carrier mobility. Yet, prior to the development of ion implantation in the early 1970s, adjustment of substrate doping was the only practical processing approach for significantly controlling V_T in active devices.

2.4.1 Ion Implantation for Adjusting Threshold Voltage

Implantation can be used either to increase or to decrease the net dopant concentration at the silicon surface. As a result, substrate doping can be selected strictly on the basis of optimum device performance since V_T can now be set by the V_T adjust implant process. In addition, since dopants can be selectively implanted into the field regions, high performance NMOS circuits can be fabricated on lightly doped substrates, without the possibility of inadvertent inversion of the surroundings field regions.

As mentioned earlier, the V_T adjust implant technique involves implantation of boron, phosphorus, or arsenic ions into the regions under the gate oxide of MOSFET. Boron implantation produces a positive shift in V_T, while phosphorus or arsenic implantation causes a negative shift. For shallow implants, the procedure has essentially the same effect as placing an additional sheet of “fixed” charge at the SiO_2-Si interface.

Ion implantation can also be used to fabricate depletion-mode MOSFETs. Depletion mode NMOS devices, in which V_T <0V are commonly used in NMOS logic circuits. In order for the required negative threshold voltage for a depletion mode NMOS device to be produced, n-type impurities are implanted into the p-type substrate to form a built in channel between the source and drain.

2.5 Subthreshold Currents in Long Channel MOSFETs

The small drain current which flows in the MOSFET channel below threshold (in weak inversion) is called subthreshold current, I_{Dst}. In most application I_{Dst} is far too small to be useful as a drive current. However, it can represent an unwanted leakage current, especially in ICs designed for low power applications. The common range of V_T in submicron digital CMOS ICs is 0.6V – 0.8V.Thus, when V_{GS} = 0V the MOSFETs in these circuit may be close to weak inversion. Consequently, for many applications, subthreshold leakage must be well characterized so that the total IC leakage current of the chip can be predicted during the design phase of the product. It should also be noted that the V_T-adjust implant generally increases the subthreshold swing St, and this is one reason that St in real devices is larger (~100Mv/dec) than the theoretically predicted value of ~60Mv/dec at 300K.

2.5.1 Subthreshold Swing, St

The subthreshold swing, St is the gate voltage change that is required for an order-of-magnitude change of the drain current in the subthreshold region. The St of a short channel is impaired by the source or drain. In the subthreshold region, the gate voltage is applied to keep the electrostatic potential of the channel sufficiently low to reduce the amount of mobile carriers in the channel and turn off the transistor. As the channel length is made sufficiently large, the subthreshold swing approaches its ideal value given by [14], i.e.,~60mV/dec at room temperature. As the channel length (L) of a typical MOSFET is reduced with all other parameters held constant, the threshold voltage decreases and the subthreshold swing, St increases. Since L decreases, the lateral fields terminate on more charge further into the channel, which essentially steals the charge that would normally be terminated by the gate voltage in a along channel devices. This stealing of charge by the lateral fields effectively lowers the source-to-channel barrier, which controls the

conduction of electrons from source to drain. As P. Vandamme et. al [15] in explaining that St depends on gate source voltage, V_{GS} and has minimum value at V_{GS} which is linearly related to the voltage at which moderate inversion starts. For large V_{GS} the drain current starts to deviate from its exponential behaviour and St is increase. According to threshold voltage roll-off and subthreshold swing rollup are commonly known as short channel effects (SCEs).In MOSFETs with uniformly doped substrates, St is calculated from

$$St = \ln 10 \, (d\ln I_D / dV_{GS})^{-1} = 2.3 \, (kT/q) \, (1+Cd/Cox) \quad \ldots\ldots (2)$$

$$= 2.3 \, (kT/q) \, [1+\kappa_{si} t_{ox}/\kappa_{ox} d)] \quad \ldots\ldots (3)$$

The factor 2.3 comes from the conversion of ln(x) to $2.3\log_{10}[x]$.

Ideally, an abrupt change in I_D should occur as V_{GS} passes through V_T. If a MOSFET exhibits a steep decline in I_D as V_{GS} is decreased below V_T, its St value will be small. Conversely, is a MOSFET structure is known to posses a small St, the implication is that only a small reduction of V_{GS} below V_T will effectively turn off the device; whereas if the device exhibits a large St value, a significantly large I_{DST} may still flow in the OFF state (when $V_{GS} = 0$).

According to equation (3), at 300K in the ideal limit of $t_{ox} = 0$, St = (2.3Kt /q) ~60mV /dec value (also, note that is T increases, so does the value of St).The St value of typical interfaces states, or a nonzero value of tox) cause St to become larger than the 60mV/dec value (also, note that if T increases, so does the value of St). Thus, if St is known, equation (2) can be used to estimate I_{Dst} in long channel MOSFETs in subthreshold operation. That is, circuit designers can readily calculate the gate bias required to keep the subthreshold negligibly small, the maximum bias applied to the gate when the devices is in the OFF state should be kept at least 0.5V below V_T.

Equation (2) also indicates that St can be made smaller by using a thinner tox or a lower substrate doping concentration which will make the channel depletion-region width, d larger. In addition, equation (2) implies that St increases with temperature. Finally, it should be noted that because the depletion width increases when a substrate bias is applied, the subthreshold swing decreases according to equation (3). **Figure 2-4** shows that the impact of short channel effects on drain current. As the channel length (L) is reduced, subthreshold swing increases (S_2>S1) and threshold voltage decreases ($V_{TH2} < V_{TH1}$) [12]

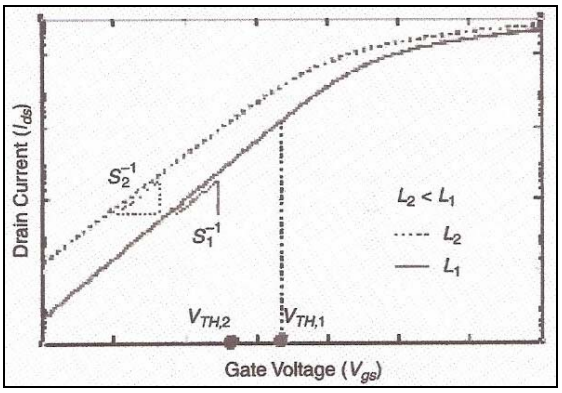

Figure 2.4: Graph of drain current, I_D vs gate voltage, V_{GS} [12]

2.5.2 Gate Induced Drain Leakage (GIDL)

Another form of leakage current observed in OFF state MOSFETs is *gate-induced drain leakage* (GIDL).The carriers responsible for GIDL originate in the region of the drain that is overlapped by the gate, and GIDL is occurs when the gate is grounded and the drain is at V_{DD}.A large electric field then exists across the oxide (ε_{ox}), which must be supported by charge in the drain region. This charge is provided by the formation of a depletion region in the drain as shown in **Figure 2.5**. If ε_{ox} becomes sufficiently large, in addition to the depletion region, an inversion

layer will attempt to form at the silicon in addition to the depletion region; an inversion layer will attempt to form at the silicon surface of the drain. However, as the minority carriers arrive at the surface to form the inversion layer, they are immediately swept laterally to the substrate. Hence in this case, the depletion region under the oxide in the drain instead becomes a deep depletion layer. The zone near the surface where an inversion layer should be formed is referred to as an "incipient inversion layer". The current that flows as a result of the carriers being swept from this incipient layer constitute the GIDL current.

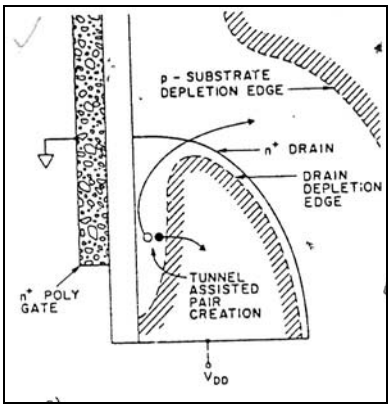

Figure 2.5: Gate induced drain leakage in the MOSFET [8]

A schematic view of the gate-drain overlap region for a grounded gate and the drain biased at V_{DD}. Tunneling created pairs lead to a lateral hole flow in the n+ drain. This flow prevents formation of an inversion layer inside the drain. Like other forms of leakage current, GIDL will contribute to standby power, giving rise to the problems of excessive heat dissipation at large device counts and heavy current drain in portable systems. Hence, GIDL must be controlled so that is does not exceed some specified maximum value, typically ≤10pA/μm.

There are several general procedures that appear to be effective in reducing GIDL. First, the oxide thickness can be increased to reduced ε_{ox} for a given voltage. But this is usually not implemented, since a thicker gate oxide will adversely impact

other devices characteristics. Second, the trap density in the near-surface area can be reduced, but such trap elimination generally requires a carefully controlled fabrication sequence. Third, the doping in the drain can be increased, as this will decrease the depletion region width and the tunneling volume. Unfortunately, the latter approaches favor abrupt junctions, because graded or lightly-doped-drain (LDD) structures will lead to a significant lateral extension of the drain within which the doping is lighter, inviting greater GIDL. Tradeoffs of field reduction against GIDL may need to be considered. On the other hand, it has also been reported that LDD devices can suppress GIDL by suppressing the lateral field.

2.6 The submicron MOSFET

Since the quest for higher density still requires L and Z to be further reduced, it will nevertheless be necessary to confront the other short channel effects. Increased "off-state" leakage in short channel MOSFET is due to several phenomena including, lowering of the threshold voltage V_T as L is decreased and/or V_{DS} is increased, the onset of punchthrough at smaller drain biases as L is decreased, and an increase in isolation leakage current as the isolation spacing is decreased. The reliability problems that arise in short MOSFETs include this gate oxide breakdown, device degradation due to hot carrier effects; and reliability problems associated with the interconnects between MOSFETs, such as electromigration failures in the metal lines.

2.6.1 Comparison of Long Channel and Short Channel MOSFET Characteristics

As a result, short channel device effects can be correlated with reduction in the gate length and/or gate width dimension. Long channel device characteristics

that undergo variation as the gate dimensions are decreased. From the comparison above, we can see that the short channel effects can be divided into the following categories: (a) those that impact $V_{T,\,(b)}$ those that impact subthreshold currents; and (c) those that impact I_D when the MOSFET is operated in saturation, $V_{DS} > (V_{GS} - V_T)$.

Long Channel MOSFET Behavior

1. The threshold voltage, V_T is independent of channel length L and width Z.
2. V_T is independent of drain bias voltage.
3. V_T depends on V_{BS}
4. The subthreshold current I_{Dsat} increases linearly as L decreases.
5. I_{Dst} is independent of drain bias.
6. The subthreshold swing St is independent of gate length.
7. The drain current in saturation I_{Dsat} is independent of V_{DS}.
8. I_{Dsat} is proportional to $(V_{GS}- V_T)^2$
9. I_{Dsat} is proportional to 1/L

Short Channel MOSFET Behavior

1. V_T decreases as L is decreased. V_T may also be impacted by changes in Z.
2. V_T decreases with increasingly V_{DS}
3. V_T increases less rapidly with V_{BS}
4. I_{Dst} increases more rapidly than linearly as L is decreases
5. I_{Dst} increases with increasing V_{DS}
6. St increases with decreasing L
7. I_{Dsat} increases as V_{DS} increases
8. I_{Dsat} is proportional to $(V_{GS} - V_T)$
9. As $L \rightarrow 0$, I_{Dsat} becomes independent of L

2.7 Punchthrough in Short Channel MOSFETS

Punchthrough is a phenomenon associated with the merging of the source and drain depletion regions in the MOSFET. That is, as the channel gets shorter, these depletion region edges get closer. When the channel length is decreased to roughly the sum of the two junction depletion widths, punchthrough is established. Nevertheless, since the depletion regions of a pn junction widen as reverse bias is increased, all MOSFETs would eventually enter punchthrough if a high enough V_{DS} could be applied. However, in MOSFETs with L>2.0μm, breakdown of the drain substrate junction generally sets in before this punchthrough voltage is reached. As a result, in practice, punchthrough is not a limiting factor in long channel digital MOSFET design. In shorter channel device, however, punchthrough does represent a serious limitation.

2.8 Short Channel Effects on the I-V Characteristics of MOSFETS Operated in the Strong Inversion Regime

Short channel effects significantly alter the dc I_D-V_{DS} characteristics of long channel MOSFETs being operated in strong inversion in three ways. First, the combined effects of reduced gate length and gate width produce a change in V_T. Second, the mobility of the carriers in the channel is reduced by two effects, which in turn reduces I_D. These two effects are the mobility degradation factor (due to the gate length), and the velocity saturation factor (due to the lateral channel field).Third, the channel length is modulated by the drain voltage when the devices is in saturation, when $V_{DS} > (V_{GS} - V_T)$, causing an increase in I_{Dsat} with increasing V_{DS} (channel modulation effect).

2.9 MOSFET Scaling

A roadmap of the scaling of MOSFET devices was proposed by Hu in 1993 [16].He assumed that a new generation of technology will continue to be developed every three years, with perhaps a slow-down to four years beyond the 0.35μm generation. His proposal for MOSFET scaling is outline as follows:

1. The IC industry/market will agree on the next power-supply voltage standard V_{CC} several years in advance of introducing a technology/ product using that voltage.
2. For a given V_{CC}, the thinnest gate oxide will be used to get maximum I_D
3. Junction depths will be scaled aggressively to keep the short channel effects within a desired limit.
4. V_T of general purpose technology will remain basically unchanged. A significant reduction in V_T is unacceptable for channel subthreshold leakage.
5. The well doping concentration or punchthrough implant dose will be increased, and the gate length may be chosen to be larger than the minimum feature size in order to achieve acceptable leakage and standby current.
6. Drain engineered structured will be used as necessary to meet the constraints of hot carrier reliability, breakdown voltage and GIDL.

The fundamental issue of downsizing MOSFETs is to preserve long-channel characteristics after miniaturization. Several approaches have been proposed as roadmaps for designing submicron MOSFETs (L <1μm) so that they exhibit such behavior. Three of the most important scaling methodologies are the constant electric field scaling and its derivatives (constant voltage and constant electrostatics scaling). The method of constant electric field scaling for designing submicron MOSFETs was proposed by Hayes [17].In this approach, a successful larger device structure is selected and all its dimensions and voltages are reduced by a constant scaling factor λ (>1).Then, scaling based on subthreshold behavior, scaling to achieve a desired value of subthreshold current, I_{off}. Because subthreshold conduction is the dominant short channel performance issue that limits MOSFETs scaling.

2.9.1 Subthreshold Swing, St

Another approach to scaling which overcomes the above difficulty was proposed by Brews et al [18].They introduced an empirical formula which determines the minimum gate length a MOSFET can have so that its subthreshold behavior remains insensitive to drain bias. That is, by using this formula a constraint upon each single combination of device parameters is identified, such that if this constraint long channel subthreshold behavior. When long channel devices are operated in subthreshold, I_{Dst} is independent of drain to source voltage once V_{DS} exceeds a few kT/q [19].Thus, the criterion selected to represent acceptable long channel behavior was not more than a 10% change in drain current could occur for a 0.5V change in V_{DS}.

2.9.2 Subthreshold Scaling

When long channel devices are operated in subthreshold, I_{Dst} is independent of drain to source voltage once V_{DS} exceeds a few kT/q. Thus, the criterion selected to represent acceptable long channel behavior was that no more than a 10% change in drain current could occur for a 0.5V change in $V_{DS.}$